This is an open access article published under a Creative Commons Non-Commercial No Derivative Works (CC-BY-NC-ND) Attribution License, which permits copying and redistribution of the article, and creation of adaptations, all for non-commercial purposes.

Environmental Science & Technology

Article

pubs.acs.org/est

Land Use Regression Models for Ultrafine Particles in Six European Areas

Erik van Nunen,[*,†] Roel Vermeulen,[†] Ming-Yi Tsai,[‡,§,∥] Nicole Probst-Hensch,[‡,§] Alex Ineichen,[‡,§] Mark Davey,[‡,§] Medea Imboden,[‡,§] Regina Ducret-Stich,[‡,§] Alessio Naccarati,[⊥] Daniela Raffaele,[⊥] Andrea Ranzi,[#] Cristiana Ivaldi,[∇] Claudia Galassi,[○] Mark Nieuwenhuijsen,[◆,¶,∞] Ariadna Curto,[◆,¶,∞] David Donaire-Gonzalez,[◆,¶,∞] Marta Cirach,[◆,¶,∞] Leda Chatzi,[†,‡] Mariza Kampouri,[⊗] Jelle Vlaanderen,[†] Kees Meliefste,[†] Daan Buijtenhuijs,[†] Bert Brunekreef,[†] David Morley,[☆] Paolo Vineis,[⊥,☆] John Gulliver,[☆] and Gerard Hoek[†]

[†]Institute for Risk Assessment Sciences (IRAS), division of Environmental Epidemiology (EEPI), Utrecht University, Utrecht, The Netherlands

[‡]Swiss Tropical and Public Health (TPH) Institute, University of Basel, Basel, Switzerland

[§]University of Basel, Basel, Switzerland

[∥]Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, Washington United States

[⊥]Human Genetics Foundation, Turin, Italy

[#]Environmental Health Reference Centre, Regional Agency for Prevention, Environment and Energy of Emilia-Romagna, Modena, Italy

[∇]ARPA Piemonte, Turin, Italy

[○]Unit of Cancer Epidemiology, Citta' della Salute e della Scienza University Hospital and Centre for Cancer Prevention, Turin, Italy

[◆]ISGlobal, Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain

[¶]Department of Experimental and Health Sciences, Pompeu Fabra University (UPF), Barcelona, Spain

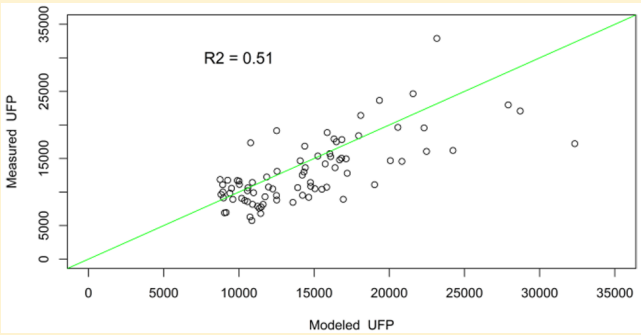
[∞]CIBER Epidemiologia y Salud Pública (CIBERESP), Barcelona, Spain

[⊗]Department of Social Medicine, University of Crete, Heraklion, Greece

[☆]MRC-PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, Imperial College London, St Mary's Campus, London, United Kingdom

Supporting Information

ABSTRACT: Long-term ultrafine particle (UFP) exposure estimates at a fine spatial scale are needed for epidemiological studies. Land use regression (LUR) models were developed and evaluated for six European areas based on repeated 30 min monitoring following standardized protocols. In each area; Basel (Switzerland), Heraklion (Greece), Amsterdam, Maastricht, and Utrecht ("The Netherlands"), Norwich (United Kingdom), Sabadell (Spain), and Turin (Italy), 160−240 sites were monitored to develop LUR models by supervised stepwise selection of GIS predictors. For each area and all areas combined, 10 models were developed in stratified random selections of 90% of sites. UFP prediction robustness was evaluated with the intraclass correlation coefficient (ICC) at 31−50 external sites per area. Models from Basel and The Netherlands were validated against repeated 24 h outdoor measurements. Structure and model R^2 of local models were similar within, but varied between areas (e.g., 38−43% Turin; 25−31% Sabadell). Robustness of predictions within areas was high (ICC 0.73−0.98). External validation R^2 was 53% in Basel and 50% in The Netherlands. Combined area models were robust (ICC 0.93−1.00) and explained UFP variation almost equally well as local models. In conclusion, robust UFP LUR models could be developed on short-term monitoring, explaining around 50% of spatial variance in longer-term measurements.



INTRODUCTION

Numerous studies have shown associations of particulate matter air pollution characterized as particles smaller than 10 μm (PM_{10}) or 2.5 μm ($PM_{2.5}$) and adverse health effects.[1,2] Much less is

Received: November 23, 2016
Revised: February 26, 2017
Accepted: February 28, 2017
Published: February 28, 2017

 © 2017 American Chemical Society

DOI: 10.1021/acs.est.6b05920
Environ. Sci. Technol. 2017, 51, 3336−3345

known about health effects of particles smaller than 0.1 μm, also known as ultrafine particles (UFP), which may be more toxic because of their potential to penetrate deeper into the lungs, their higher biological reactivity per surface area, and their potential uptake in the bloodstream.[3,4] UFP contributes only a small fraction to particle mass and thus UFP is not well reflected by PM_{10} or $PM_{2.5}$ measurements.[5] The lack of data on health effects of long-term UFP exposure is related to a lack of routine monitoring and models describing the large spatial variation of UFP.[5] Therefore, there is a need for models that provide long-term UFP exposure estimates at a fine spatial scale.

Land Use Regression (LUR) models are a common approach in epidemiology to assess air pollution exposure at a fine spatial scale, using predictor variables from Geographic Information Systems (GIS). LUR models for $PM_{2.5}$ and NO_2 are typically built on data from (bi)weekly measurements at 20−80 monitoring sites per study area.[6,7] Few studies applied this monitoring strategy to UFP.[8,9] However, because of high costs and labor-intensive operation of UFP monitors, this approach is not attractive for UFP. Recent studies developed UFP LUR models based on short-term monitoring[10−15] or mobile monitoring campaigns conducted while driving.[15−18] Previously published short-term and mobile UFP models substantially differed in model structure (GIS predictors included in the model) and model performance (percentage explained variability (R^2)). Due to differences in area size, number of monitoring sites, duration, and frequency of monitoring, monitoring equipment, GIS predictor variables, and model development procedures, it is unclear whether the difference in model structure and performance is due to inherent differences between study areas or due to these methodological issues. A recent study showed that models based on short-term and mobile monitoring in the same study area resulted in comparable model structures and highly correlated predictions at external sites.[15]

Most studies develop a single best model, which is applied for exposure assessment in epidemiological studies. Due to correlations between predictor variables, it is likely that alternative models can be developed which explain variability almost equally well.[12] Gulliver[19] developed and interpreted four NO_2 models in the framework of 4-fold hold-out validation (HV). Wang[20] applied model predictions of 40 models from a cross-validation method to predict subject's exposure to NO_2 in an epidemiological study. Very few studies have developed multiple models for short-term monitoring designs (Hankey, 2015). Little is known about the robustness of model predictions at external sites by applying multiple models developed on one monitoring data set. Using external sites is important as for short-term and mobile monitoring, the monitoring sites used for model development may differ systematically from the often residential addresses to which the model are applied, for example, in distance to roads.

We performed a harmonized short-term monitoring campaign contemporaneously in six European study areas. We developed ten LUR models per area based on 90% subsets of the sites, following a common modeling approach. Our aims were to develop LUR models for predicting spatial patterns in UFP for six European study areas; to assess the agreement in LUR model structure and performance within and between study areas. A further aim was to evaluate the performance of a model using the UFP concentration data from six study areas combined. Important new contributions of this paper include (a) the evaluation of the robustness of model predictions at external residential sites, not included in model development in all six areas; (b) Validation of the models with UFP monitoring data with longer monitoring duration at residential external sites in two of the areas; (c) an evaluation of the potential to develop a model for a large geographic area and comparison with performance of local models.

■ MATERIALS AND METHODS

Study Areas and Design. In Basel (Switzerland), Heraklion (Greece), Amsterdam, Maastricht, and Utrecht (The Netherlands, three cities collectively referred to as "The Netherlands"), Norwich (United Kingdom), Sabadell (Spain), and Turin (Italy), monitoring sites were selected based on criteria applied before in the ESCAPE and MUSiC studies,[7,12,21] and evaluated by a team of experts from all centers (Supporting Information (SI) 1). In each area, 160 sites were selected (240 in The Netherlands because multiple cities were studied). For large spatial contrast in traffic intensities and land use, seven types of monitoring site were defined: traffic, urban background, urban green, water, highway, industry, and regional background, as applied before.[21] Measurements were made as close as possible to home façades, but not on private property. Traffic sites were monitored close to home façades along a major road with >10 000 vehicles/day, not on curbsides. Urban background sites were close to home façades >100m away from a major road. Urban green sites were at the edge of a park, water sites adjacent to a canal or a river, highway sites were within 100m from a highway, industry sites were in a mixed industrial-residential zone, and regional background sites were outside the study city. Traffic sites represented approximately 40% of the total sites in all areas.

In all areas, a harmonized short-term monitoring campaign was conducted contemporaneously between January 2014 and February 2015, measuring each monitoring site three times in different seasons (Summer, Winter and Spring/Autumn). Measurements were taken on Monday-Friday, and site types were visited in random order. At each visit, UFP concentrations were measured for 30 min following a prescribed protocol, and a GPS coordinate was taken. To avoid rush hour influences and increase comparability between monitoring sites, measurements were taken between 9.00 am and 4.00 pm. During the entire measurement campaign, reference site UFP measurements were conducted in each area to allow temporal adjustment of local data. The reference site was an urban background location in the study area (SI 1). In the large study area of The Netherlands, the reference site was in one of the areas (Utrecht), 40 km from Amsterdam and 140 km from Maastricht.

Monitors. UFP was monitored in all study areas using a CPC 3007 (TSI Inc., Shoreview, MN), operating at a flow of 100 mL/min measuring particles ranging from 10−1000 nm at 1 s intervals. The CPC 3007 does not specifically measure UFP, but UFP typically dominates particle number.[5] We will use the term UFP to refer to the particle number counts. The reference sites in The Netherlands and in Heraklion were also equipped with a CPC, operating at identical settings, whereas other areas used a MiniDiSC (Testo AG, Lenzkirch, Germany), because of the limited number of CPCs available. The MiniDiSC operated at a flow of 1000 mL/min measuring particles from 10−300 nm at 1 s intervals. Previous studies had shown good agreement between CPC 3007 and MiniDiSC.[22,23] We colocated the two instruments used in each study area regularly to check comparability. In The Netherlands, Norwich and Sabadell, the mean ratio of the two instruments was close to unity (SI 2). In Turin, the CPC used at the short-term sites gave 27% lower readings than the MiniDiSC used at the reference site.

DOI: 10.1021/acs.est.6b05920
Environ. Sci. Technol. 2017, 51, 3336−3345

In Heraklion, the monitoring site CPC gave 41% higher UFP readings than the reference site CPC with large variation. We did not correct for these differences, as the reference site measurements is used only to correct for temporal variation. GPS coordinates were collected using a high sensitivity hand-held GPS device.

Data Cleaning. QA/QC included zero checks before and after measurements and regular colocation of all UFP monitors per study area at the local reference site for at least 3 h per exercise. All site and reference measurements were averaged over the corresponding period, after removing measurements with error codes of the instrument (e.g., deviating flow). Extreme reference site 30 min measurements, defined as more than four interquartile ranges (IQRs) lower or higher than the 25th or 75th percentile, were flagged and individually inspected, as they might indicate local sources near the reference site (e.g., diesel-powered grass mower near the Dutch site) not reflective of concentration patterns in the wider area. We identified 15, 30 min reference observations as indicative of local sources (3 in The Netherlands, 10 in Norwich, 2 in Sabadell), 0.5% of all reference site observations.

In Turin, reference site measurements were missing for 65% of the 480, 30 min measurement periods due to misinterpretation of the protocol. A regression model using Routine NO_x, Hour of the day, Barometric pressure and Relative Humidity, fit on the valid 35% of the data (R^2 62%), was applied to impute the missing 30 min reference site observations (SI 3). In Norwich 17% of the reference site data was missing due to operational problems. A regression model built on routine and meteorological data (R^2 50%) was used to impute these missing observations (SI 3). In the other areas, no predictive model could be developed (percentage missing <10% in Netherlands, Basel, and Sabadell and 18% in Heraklion).

Temporal Variation Adjustment. To improve assessment of spatial contrasts between sites, the UFP concentration at the local reference site was used to adjust monitored UFP levels for temporal variability in three steps, following procedures of previous studies.[12,24] First, the mean reference UFP concentration of the corresponding interval was subtracted from the annual mean concentration at the reference site. Second, this difference was added to the concentration measured at a site. Third, the adjusted average UFP concentration was calculated as the average of three adjusted samples from one site. Application of the ratio method (accounting for differences between two instruments) resulted in unrealistic averages due to large individual ratios (up to 8) on days with low UFP concentrations at the reference site.

GIS Predictors. GPS coordinates from three site visits were averaged and manually corrected for optimal accuracy in position relative to roads on detailed road maps. Predictor variables were generated locally for each of these sites in a GIS, using coordinates and digital data sets on traffic, heavy traffic, population density, land use and restaurant density. Predictors and buffer sizes were similar to these used in the ESCAPE and MUSiC studies,[7,12] supplemented with airport land use and restaurant data because of studies documenting increased outdoor UFP concentrations related to emissions from airports[25,26] and restaurants,[27] and the inclusion of restaurants in a previous UFP model.[11] Traffic and heavy traffic predictors were collected at buffer radii of 50, 100, 300, 500, and 1000 m from the best available road network data (SI 4). Population and land use predictors at radii of 100, 300, 500, 1000, and 5000 m (Land use defined as airport only radii of 1000 and 5000 m) were collected from population density data from the European Environmental Agency and CORINE land use data sets (Coordination of Information on the Environment). Number of restaurants was collected at radii of 100, 300, 500, 1000, and 5000 m using the Open Street Map application Turbo Overpass. Heavy traffic data from Basel, Heraklion, Sabadell and Turin were not available in a GIS. Restaurant data do not cover all restaurants in the city as inclusion in the database is not free (SI 4). Restaurant data were not used for Heraklion, since the number of amenities was underreported and did not reflect realistic distributions across neighborhoods.

External Sites. We used external sites to test the robustness of predictions of the 10 LUR models. Residential addresses of 31−48 subjects per study area participating in the EXPOsOMICS study[28] were used for all areas except Heraklion. In Heraklion, 50 randomly selected addresses were used. GIS predictors for subject's home addresses were collected to test robustness of model predictions. Additionally, in Basel and The Netherlands 24 h average outdoor UFP concentrations were monitored at the home façade with MiniDiSCs in three seasons. Study period and study area were harmonized between the short-term monitoring campaign and residential outdoor measurements. The temporally adjusted average UFP concentration was used for external model validation when at least two valid 24 h observations were available.

LUR Model Development. LUR models were developed centrally by applying procedures equivalent to procedures applied in the ESCAPE and MUSiC studies.[7,12] Briefly, temporal-variation adjusted 30 min average UFP concentration per site was used as dependent variable in a linear regression model, using GIS predictors as explanatory variables. Predictors where the 90th percentile was zero were not used in any model. Predictors that were not available for all areas or present in less than 50% of the areas were not used in the combined area model. Predictors were selected using a supervised stepwise selection procedure, selecting the variable with the largest adjusted R^2 to the model if the direction of effect was as defined a priori and did not change the direction of effect of previously included variables. This process was continued until no more variables provided a gain in adjusted R^2. Variables included were checked for p-values (removed when p-value > 0.10), collinearity (removed when variance inflation factor > 3), and influential observations (if Cook's D > 1 the model was further examined).

Local UFP Models. In each area, 10 models were developed to evaluate robustness of model structure and model predictions at the external sites, following the 10-fold cross-validation approach. First, monitoring sites were stratified by site type (traffic vs nontraffic) and subsequently randomly distributed in 10 groups. Next, each time 90% (9 groups) of the sites was used for model development and 10% (1 group) for validation. The model R^2 and root mean square error (RMSE) were obtained from each individual model, the HV R^2 and RMSE were obtained by predicting UFP levels in each validation set and regressing these against measured values over all pooled random draws. In Basel and The Netherlands, an additional validation was obtained by testing modeled against measured 24 h outdoor UFP concentrations at the external sites. We calculated bias, defined as the average of modeled minus measured UFP.

For model structure comparison, we classified predictors in nearby traffic (traffic predictors, radius ≤100m), distant traffic (traffic predictors, radius > 100m), population, industry, port, airport, restaurants, and green space. Predictions from the 10 models at external sites in a specific area were compared with

DOI: 10.1021/acs.est.6b05920
Environ. Sci. Technol. 2017, 51, 3336−3345

scatterplots and correlation coefficients. The intraclass correlation coefficient (ICC) was calculated as a summary. Predictions were performed after truncation of predictors such that they were within the range in the model development data (truncation applied on one site in Basel, two sites in Heraklion, two sites in The Netherlands, three sites in Norwich). A chart of procedures is presented in Figure 1.

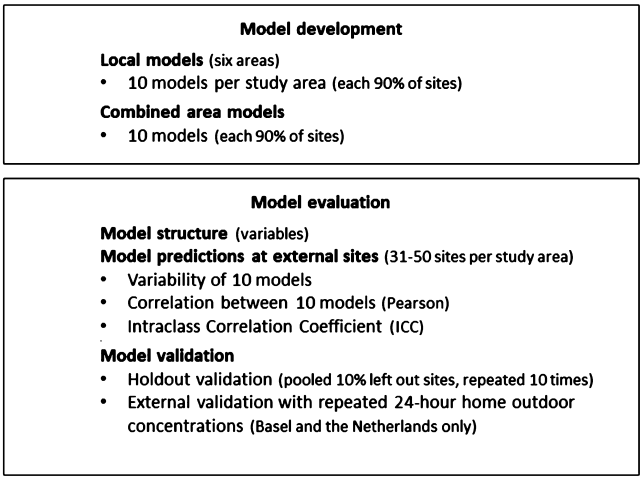


Figure 1. ; Overview of model development and evaluation. External sites are residential addresses of participants in the personal exposure monitoring survey.

Combined Area UFP Models. Ten models on combining data from all areas were developed following the procedure described before, additionally stratifying sites by study area prior to stratification by site type. To account for systematic differences in background concentration between study areas, we specified random intercepts using a linear mixed-effect model after the supervised stepwise model development procedure. We further evaluated random slopes to account for differences in emissions due to, for example, composition of the vehicle fleet across areas.

Leave one area out validation (LOAOV) was applied to explore applicability of combined LUR models in areas without measurements. All short-term sites from one area were excluded and one model was developed for all other areas. A random intercept per area was introduced and the LOAOV R^2 and RMSE were obtained by evaluating modeled and measured UFP levels in the excluded area. For Basel and The Netherlands LOAOV models were also compared with measurements at the external sites.

GIS predictors were generated locally in ArcGIS (ESRI, Redlands, CA) (land use, population, and traffic predictors) and in the Overpass Turbo[29] and QGIS[30] applications (Restaurant predictors). Local data cleaning and calculations per center were performed using the statistical package available (SAS, STATA, R), final checks and model building were performed using the statistical package R 3.2.2.[31]

RESULTS

For LUR model development, 160 monitoring sites per city in Basel, Heraklion, Sabadell and Turin, 161 sites in Norwich and 242 sites in The Netherlands were monitored (total 1043 sites). Adjusted average UFP observation were included for LUR modeling when based on at least two 30 min site observations, corrected for corresponding reference measurements, leading to loss of 1 site in Basel, 2 sites in The Netherlands and 10 sites in Heraklion, an overall loss of 13 sites (1.2%). There was large variability in adjusted average UFP concentrations among sites in all study areas (Figure 2). Concentrations were highest at the traffic sites and industrial sites in Turin and Sabadell. Higher median UFP concentrations were observed in Sabadell and Turin. Variability of the individual three 30 min observations was high. The average within site standard deviation after temporal adjustment was 6985 particles/cm^3, 51% of the overall mean across study areas.

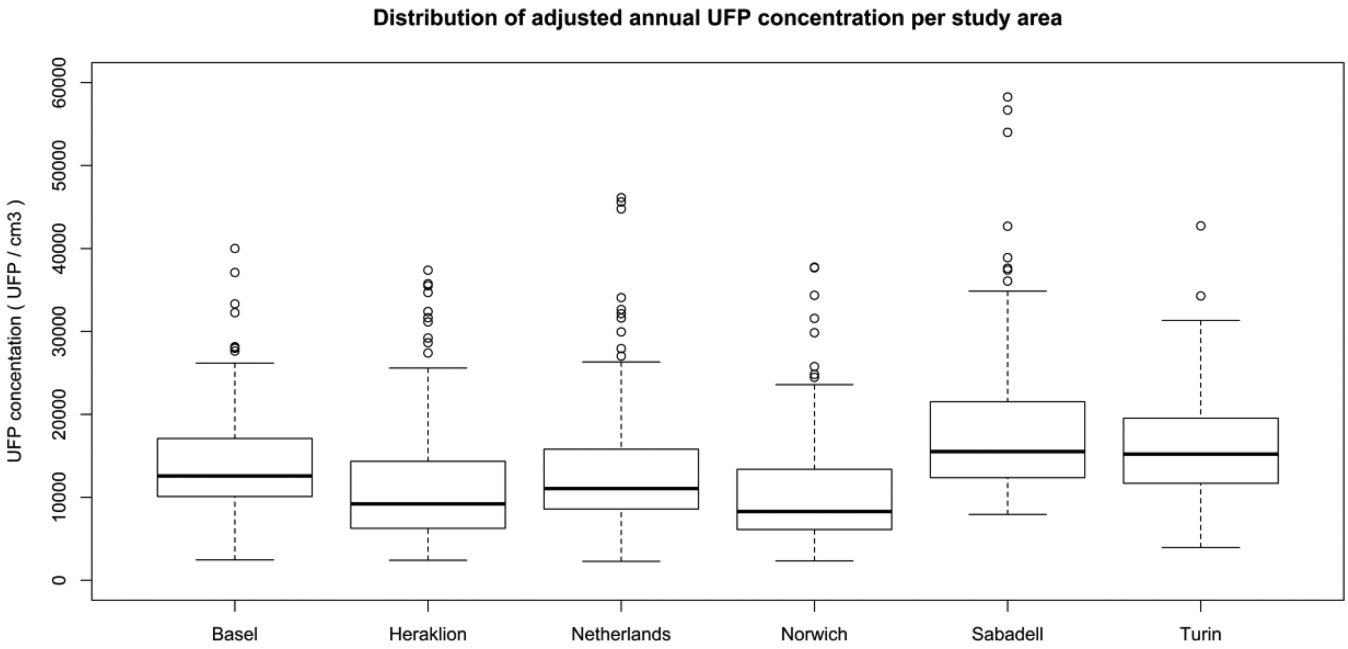


Figure 2. Distribution of average UFP concentrations (cm^{-3}) per study area. The Netherlands is comprised out of the cities of Amsterdam, Utrecht, and Maastricht.

DOI: 10.1021/acs.est.6b05920
Environ. Sci. Technol. 2017, 51, 3336−3345

Table 1. Model Performance and Robustness of Prediction of Local LUR Models Sites[a]

local models		Basel	Heraklion	Netherlands	Norwich	Sabadell	Turin
model performance		***n* = 159**	***n* = 150**	***n* = 240**	***n* = 161**	***n* = 160**	***n* = 160**
model R^2 (%)	mean (SD)	30 (2)	37 (4)	48 (2)	39 (2)	28 (3)	40 (2)
model RMSE (UFP/cm^3)	mean (SD)	5251 (175)	6128 (220)	5511 (169)	5149 (132)	7507 (354)	4676 (126)
HV R^2 (%)		18	17	35	25	18	33
HV RMSE (UFP/cm^3)		5611	6930	5548	5672	8247	4913
HV bias (UFP/cm^3)		−25	−49	37	−74	88	81
		***n* = 40**		***n* = 41**			
external R^2 (%)	mean (SD)	54 (2)		49 (5)			
external RMSE (UFP/cm^3)	mean (SD)	2827 (70)		3524 (118)			
external bias (UFP/cm^3)	mean (SD)	2675 (291)		471 (223)			
model robustness		***n* = 48**	***n* = 50**	***n* = 42**	***n* = 31**	***n* = 42**	***n* = 44**
UFP/cm^3	mean (SD)	13625 (291)	11131 (931)	15414 (223)	9565 (285)	17752 (225)	14714 (99)
ICC		0.98	0.73	0.97	0.86	0.96	0.98

[a]Explained variability (R^2), root mean square error (RMSE), and bias (difference between modeled and measured UFP) from the 10 models at model development, holdout validation (HV, based on pooled analysis), and at application on external sites. Model robustness expressed in mean and SD of predicted UFP/cm^3 and intraclass correlation coefficient (ICC) of 10 models at application on external sites.

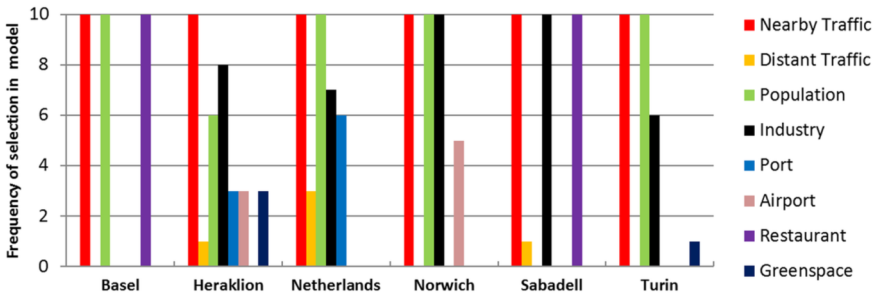


Figure 3. Selection of predictor categories in the 10 models per study area.

Local LUR Models. Model R^2 differed between areas, ranging from on average 28% in Sabadell to 48% in The Netherlands. Model R^2 and RMSE of the ten models within areas were very similar (Table 1). Within an area, the 10 models typically contained one to three predictor categories (e.g., nearby traffic) in all 10 models (Figure 3). Other predictor variables were included in a selection of models, such as port -included in 6 of 10 models in The Netherlands- or industry which was included in 6 of 10 models in Turin. The exact predictor variables (e.g., traffic intensity nearest road) and coefficients differed more among the 10 models (SI 5). Between study areas more difference in model structures was seen (Figure 3, SI 5). Nearby traffic was included in all models, population was included in 46 of 60 models (not at all in Sabadell), industry was included in 41 of 60 models (not at all in Basel), and restaurant data were included in all local models in Basel and Sabadell, but not in any model of the other study areas.

HV R^2 decreased by 7−20% compared to model R^2 and RMSE increased by about 10% (Table 1). The models predicted UFP variability at external sites with longer duration monitoring substantially better (R^2 in Basel 53% and in The Netherlands 50%) At the external sites, there was virtually no bias for The Netherlands and a modest 20% systematic overestimation at the Basel sites.

Consistent with the modest differences in local model structure, UFP predictions among models per area were highly correlated (Figure 4, Table1). Predictions in individual models showed high similarity in Basel, The Netherlands, Sabadell, and Turin (ICC 0.96−0.98) and more variation in Norwich and especially Heraklion (Figure 4 and SI 6). Because of the high consistency of models, we also developed models based on 100% of the sites (SI 7). In each area, models were very similar to the 10 models per area.

Combined area LUR Models. Final LUR models included a random intercept for study area. A random slope per area did not improve prediction and was not included (SI 5 and 8). Models built on short-term sites from all areas resulted in a Model R^2 of 34% with low SD (Table 2). Every model consisted of predictors representing nearby traffic, distant traffic, population and industry (SI 5). Modeled concentrations of the 10 models on external sites were highly correlated (Table 2 and SI 6). HV R^2 over all areas was close to the model R^2. HV R^2 and RMSE of the combined model assessed per area were similar to HV R^2 and RMSE of local models (Table 3). Validation R^2 at external sites in both Basel and The Netherlands was higher than HV R^2, comparable to performance in local models.

We further tested the combined model by dropping complete areas from the model development (SI 9). The LOAOV R^2 was close to the HV R^2 of local and combined models. When applying LOAOV models on external sites, it performed equally well as local and combined area models in Basel, where in The Netherlands R^2 decreased and RMSE increased (Table 3). Systematic overestimation (Heraklion, Turin) and underestimation (Sabadell, Switzerland) up to about 30% of the overall mean were found for combined models excluding complete areas. At external sites overestimation of about 2% (Netherlands) and 20% (Basel) were found. Measurements at the external sites were 24 h averages, including night-time with typically lower concentrations.

DOI: 10.1021/acs.est.6b05920
Environ. Sci. Technol. 2017, 51, 3336−3345

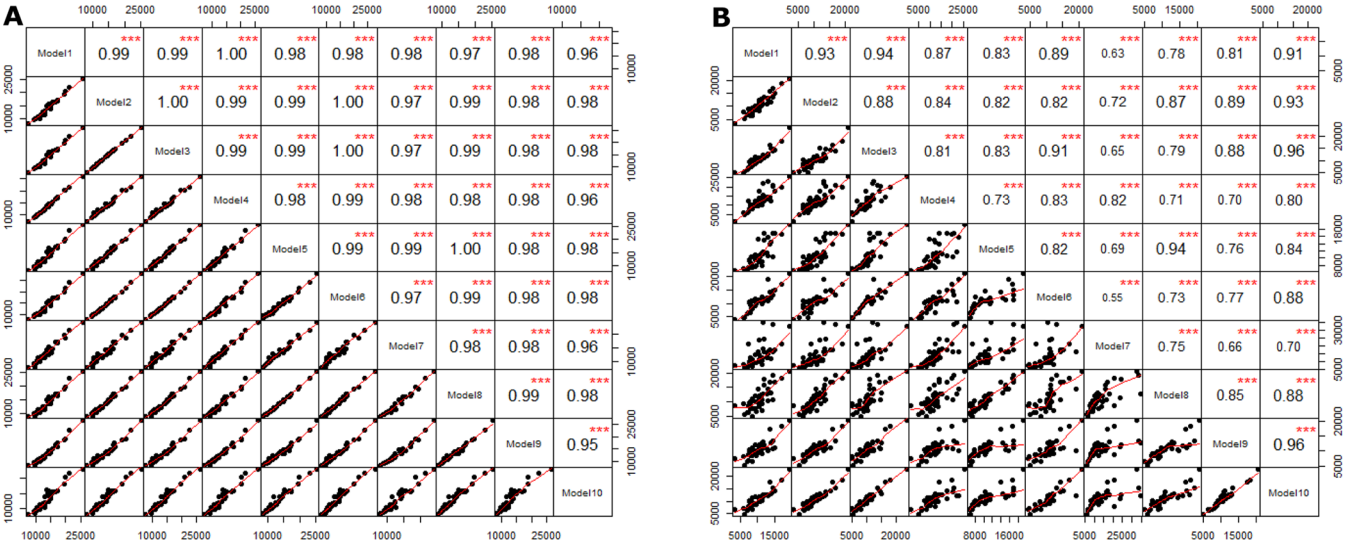


Figure 4. ; Predicted UFP concentrations of each model plotted against each other for Basel (A, highly similar) and Heraklion (B, more variation) in the lower panel, together with the Pearson correlation coefficient in the upper panel. Red lines represent the best fit lines; *** = p-value < 0.001.

Table 2. Model Performance and Robustness of Prediction of Combined Area LUR Models[a]

combined models		Basel	Heraklion	Netherlands	Norwich	Sabadell	Turin
				n = 1030			
model performance		n = 159	n = 150	n = 240	n = 161	n = 160	n = 160
model R^2 (%)	mean * (SD)			34 (1)			
model RMSE (UFP/cm^3)	mean * (SD)			6105 (117)			
HV R^2 (%)	pooled			32			
	per area	18	15	38	26	18	27
HV RMSE (UFP/cm^3)	pooled			6170			
	per area	5615	7012	5403	5631	7963	5149
HV bias (UFP/cm^3)	pooled			1			
	per area	205	120	59	106	−554	−184
		n = 40		n = 41			
external R^2 (%)	mean (SD)	52 (1)		51 (1)			
external RMSE (UFP/cm^3)	mean (SD)	2827 (25)		3466 (49)			
external bias (UFP/cm^3)	mean (SD)	2433 (289)		790 (162)			
model robustness		n = 48	n = 50	n = 42	n = 31	n = 42	n = 44
UFP/cm^3	mean SD	13 351 (289)	11 760 (281)	15 722 (162)	10 446 (118)	18 388 (334)	15 720 (339)
ICC		0.99	0.93	1.00	0.99	0.99	1.00

[a]Explained variability (R^2), root mean square error (RMSE), and bias (difference between modeled and measured UFP) from the 10 models at model development, holdout validation (HV, based on pooled analysis), and at application on external sites. Model robustness expressed in mean and SD of predicted UFP/cm^3 and intraclass correlation coefficient (ICC) of 10 models at application on external sites. [b]Based on values prior to introduction of Random Intercept

DISCUSSION

LUR models for UFP were developed in six European areas based on harmonized short-term monitoring campaigns and a common modeling approach. The 10 models developed within each area were generally robust in model structure and in prediction at external sites. Model structure differed between the six areas. Model and HV R^2 were low to moderate. Validation at external sites with repeated 24 h monitoring in two of the six areas showed substantially higher R^2s (50−53%). A combined area model explained UFP variability at external sites from two areas equally well as local models.

Robustness of LUR Model Predictions within Areas. Predictor categories selected in the 10 models per area had high agreement, resulting in highly correlated model predictions at external sites. Exact predictors in final models could differ, but due to correlation of predictors within a predictor category, modeled UFP concentrations were highly correlated. Variables like traffic intensity and heavy traffic intensity on the nearest road, variables of two adjacent buffer categories (e.g., 300 and 500 m) as well as population and address density, were correlated as observed before.[12] Predicted UFP levels from local models were very consistent in four of the areas, with slightly higher variability in Heraklion and Norwich. In these two areas more moderate correlations were found with 2 of the 10 models which included the predictor traffic intensity divided by distance. In Heraklion, one of the models with lower correlation had a lower coefficient for traffic on the nearest road (the main predictor in the Heraklion models) compared to the other nine models. This likely contributed to the more modest correlation with other models.

Local LUR Model. Despite harmonized monitoring and modeling approaches, differences in model R^2, RMSE and struc-

DOI: 10.1021/acs.est.6b05920
Environ. Sci. Technol. 2017, 51, 3336−3345

Table 3. Model Performance of the Combined Area Model by Leave One Area Out Validation[a]

		Basel	Heraklion	Netherlands	Norwich	Sabadell	Turin
short-term sites		***n* = 159**	***n* = 150**	***n* = 240**	***n* = 161**	***n* = 160**	***n* = 160**
HV R^2 (%)	LOAOV	20	14	28	22	14	28
HV R^2 (%)	local	18	17	35	25	18	33
HV R^2 (%)	combined	18	15	38	26	18	27
HV RMSE (UFP/cm^3)	LOAOV	6180	7020	5700	5840	7990	5080
HV RMSE (UFP/cm^3)	local	5611	6930	5548	5672	8247	4913
HV RMSE (UFP/cm^3)	combined	5615	7012	5403	5631	7963	5149
HV bias (UFP/cm^3)	LOAOV	−1060	2416	−85	1442	−3086	3031
HV bias (UFP/cm^3)	local	−25	−49	37	−74	88	81
HV bias (UFP/cm^3)	combined	205	120	59	106	−554	−184
external sites		***n* = 40**		***n* = 41**			
R^2 (%)	LOAOV	53		41			
R^2 (%)	local	53		50			
R^2 (%)	combined	53		51			
RMSE (UFP/cm^3)	LOAOV	2795		3831			
RMSE (UFP/cm^3)	local	2800		3485			
RMSE (UFP/cm^3)	combined	2817		3459			
bias (UFP/cm^3)	LOAOV	1845		181			
bias (UFP/cm^3)	local	2708		481			
bias (UFP/cm^3)	combined	2434		790			

[a]Leave one area out validation (LOAOV) models are based on combined model with one complete area excluded in model development, on which the model is subsequently tested. Holdout validation (HV) R^2, root mean square error (RMSE), and bias (difference between modeled and measured UFP) of local and combined area models repeated from Tables 1 and 2. Local and combined area external R^2, RMSE and bias are based on average predicted UFP concentration of the 10 models.

ture were found between the six areas, which were much larger than differences between the 10 models within an area. Models from all areas included nearby traffic—often traffic intensity at the nearest street-, consistent with the major influence of motorized traffic emissions on urban UFP concentrations.[5] Nearby traffic variables predicted a substantial contrast in UFP, of typically 4000−6000 particles/cm^3 for a difference between the 10th and 90th percentile of the predictor. The relatively high number of traffic predictors offered is another potential explanation, however the inclusion of many more near compared to distant traffic predictors argues in favor of the source interpretation. Population density was included in all 10 models in four of the six areas, 6 out of 10 models in Heraklion, and in none of the models in Sabadell. This is possibly due to the lower population variability in this moderate sized town. Industry, port, airport and restaurants were included in models of only one or a few areas. Port in a 5 km buffer was only represented in Heraklion and The Netherlands, not located within this radius in the other areas. Airport was not selected in The Netherlands, probably because few sites were located within a 5 km radius of an airport. The inclusion of these nontraffic sources is consistent with studies documenting that UFP emissions are related to multiple combustion sources.[5]

We do not have a clear explanation of the difference in model R^2 between the six study areas. Differences in model R^2 could be due to the characteristics of the study area such as size and complexity, but also to differences in the variability of GIS predictor variables. Different performance of our temporal adjustment may have contributed to variability in model R^2 as well. In Norwich and especially Turin, imputation of measurements at the reference site was used to avoid missing values. This may have reduced the effectiveness of temporal adjustment.

The current local model R^2s, ranging from 28% to 48%, and predictors used in these models are comparable to those reported of spatial LUR models in previous short-term monitoring work. In Girona province, Spain, a model with only traffic predictors captured 36% of UFP variability at 644 sites measured for a single 15 min period.[10] For Vancouver, Canada, a single measurement at 80 locations resulted in model R^2s from 29 to 53% including traffic population, port and restaurant predictors.[11] In Amsterdam and Rotterdam, The Netherlands, 37% of UFP variability at 160 sites was explained with traffic, population and port predictors.[12]

Our spatial models can be applied for assessing long-term average exposures. We did not develop spatiotemporal models, further including temporal predictors such as temperature, to allow temporally more refined estimates.

Model Validation. HV R^2s were low to moderate in all areas of our study. A low HV R^2, however, does not imply that models do not provide valid predictions, as argued previously.[12] Current UFP models predicted repeated 24 h measurements from Basel and The Netherlands substantially better than the HV R^2 suggested. For both areas a moderately high R^2 of around 50% was found, compared to HV R^2 of 18 and 35% in Basel and The Netherlands. We previously documented higher external validation R^2 related to longer averaging times at the external validation sites relative to the model development sites in two studies.[12,15] Our spatial predictors are constant in time and therefore cannot explain remaining temporal variation in short-term measurements. Repeated 24 h measurements likely reflect long-term average UFP concentrations better than short-term monitoring, because these observations are less affected by temporal exposure variation. Model R^2 and HV R^2 from short-term monitoring may not be the metric that should be leading in assessing model performance. Based on these metrics models from The Netherlands (R^2 = 48%) were better than Basel models (R^2 = 30%), but at external sites models performed equally well. This suggests that testing on external sites with longer-term monitoring is a better tool to assess performance. Long-term

DOI: 10.1021/acs.est.6b05920
Environ. Sci. Technol. 2017, 51, 3336−3345

UFP concentration data are however not routinely available and thus require a dedicated monitoring effort, as illustrated in a recent Swiss study where external validation from routine monitoring was available for four sites for UFP and 80−100 sites for PM_{10} and NO_2.[9]

Model and HV R^2 were lower in our and most other short-term and mobile LUR models for UFP compared to LUR models developed for pollutants such as NO_2 and $PM_{2.5}$.[6,7] The large spatial variation of UFP may be more difficult to model, but the use of short-term averages for UFP compared to much longer average times for NO_2 and $PM_{2.5}$ likely explains part of the difference in model R^2. In a Swiss study, based upon 2 week monitoring periods, model R^2 was similar for UFP and $PM_{2.5}$ absorbance and higher than for $PM_{2.5}$ and NO_2.[9]

Combined LUR Models. Combined model R^2 was 34% with very high consistency across the 10 models, almost similar pooled HV R^2, and identical predictor categories represented. Within the different areas combined models performed almost similar to Local models in HV R^2 and RMSE. The relatively modest differences in UFP concentrations across study areas and the dominance of traffic as the major predictor may have contributed to the possibility to develop combined models that were only slightly less predictive than the local models. While UFP concentrations were somewhat higher in Sabadell and Turin, the difference with the other areas was lower than previously reported for pollutants such as $PM_{2.5}$, NO_2 and black carbon.[7,24,32]

The rationale for developing combined area models is especially that combined area models may be applied in areas without monitoring more readily than single area models. Models for large geographical areas for other pollutants are increasingly developed[33] and our study suggests that this approach is feasible for UFP as well. Increased model validity related to using more sites[34,35] is another rationale. Problems with developing combined models include availability and comparability of predictor data and assumptions of the same effect of a specific predictor (e.g., traffic nearest road) on concentrations. For example different traffic compositions may result in different associations to traffic related predictors per area, but this was not observed in the current study (SI 8). If a predictor variable (source) is present in a few areas only, it is difficult to distinguish the influence of this source from other systematic differences between areas. In the current study, ports were absent in four study areas. We chose to exclude port as a predictor variable in combined models. We further excluded restaurant data, as data were missing in Heraklion. This potentially contributed to the lower model R^2 compared to local models from Heraklion, The Netherlands, or Turin in the current study, since UFP variability can no longer be explained by port or restaurant.

LUR models with short-term sites from one area excluded explained UFP variability in Basel equally well as local and combined models, where LOAOV R^2 remained at 53%. In The Netherlands LOAOV R^2 dropped by 10% and RMSE increased by 10% compared to local and combined models. The Netherlands study area was the only individual study area that covered a large geographical area with both large cities and smaller towns. The LOAOV model in contrast to the local model did not include 5000 m population and address density, accounting for these urbanistic related differences. These results suggest that transferability of models to independent areas is more difficult, but this could only be tested in two areas. The use of local sites in the development of LUR models seems to be beneficial for model fit at independent sites, as shown for The Netherlands.

Implications for Epidemiological Studies. We suggest to apply all the 10 models we developed to assess long-term UFP exposure in epidemiological studies and to perform 10 epidemiological analyses. This will allow assessment of the consistency of epidemiological associations obtained with these 10 different models, improving assessment of uncertainty of effect estimates beyond standard errors. The number of models could be extended, using, for example, Monte Carlo approaches.[18] Applying multiple models will likely provide consistent associations for models with high agreement, but more variation for areas with lower ICCs (Norwich and Heraklion). Alternatively, exposure could be the average of 10 models (for example by Bayesian model averaging) applied at cohort addresses. Exposure estimates will in both cases depend less on specific selected GIS variables compared to using a single best model based on model R^2. This is particularly of interest for variables for which it is unclear whether they are causally related to UFP or are proxies for other variables. An example is the variable "port" in The Netherlands, which was selected in six of ten models. Port has been a predictor in previous UFP LUR models,[8,11,12] but in the current study could also represent other differences between the city of Amsterdam (with port) and the other two Dutch cities without ports. The inclusion of port in 6 of 10 models may reflect the uncertainty of the importance of this variable. The lack of inclusion of port in some models was not due to too few sites with a nonzero value: 77 of the sites had a nonzero value.

For epidemiological studies within the study areas covered by monitoring, we suggest to primarily use the local models. Although our study did not show large differences in performance compared to the combined model, the inclusion of more specific predictors in the local model favors its use. The combined model could be applied as a further test of consistency of epidemiological findings. As our study areas did not cover very large metropolitan areas (London, Paris), Northern or Central and Eastern Europe, rural areas, nor altitude differences, we cannot apply the model with confidence across Europe. We therefore advise to apply the combined model in urban areas similar to the monitored areas. A combined model is furthermore more useful in multicity studies than in single city studies, particularly if between-city contrasts in exposure are exploited.[33]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.est.6b05920.

1. Study areas; 2. Co-location of UFP monitors; 3. Imputing missing Reference Site UFP concentrations; 4. GIS predictors for Land Use Regression Modeling; 5. Local and combined area LUR models; 6. Robustness of predicted UFP concentrations; 7. Models developed upon 100% of the sites; 8. Mixed-Effect Models Combined area LUR models; 9. Leave One Area Out combined models (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +31 30 253 9474; e-mail: e.vannunen@uu.nl.

ORCID

Erik van Nunen: 0000-0001-6079-0030

David Morley: 0000-0002-5092-2762

DOI: 10.1021/acs.est.6b05920
Environ. Sci. Technol. 2017, 51, 3336−3345

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was funded by the EU seventh Framework Program EXPOSOMICS Project. Grant Agreement No.: 308610, and the Compagnia di San Paolo (Turin, Italy) to Paolo Vineis. We are very grateful to the following people for their contribution: Jules Kerckhoffs, Cristina Vert Roca, Annemarie Melis, Andreas Schwärzler, Gregor Juretzko, Katja Stähli, Sandra Okorga, Benjamin Flueckiger, Lourdes Arjona, Pau Pañella, Danai Dafni, and Minas Iak. We thank Maastricht University and the municipality of Amsterdam for using their facilities during the short-term monitoring campaigns.

■ REFERENCES

(1) Brook, R. D.; Rajagopalan, S.; Pope, C. A.; Brook, J. R.; Bhatnagar, A.; Diez-Roux, A. V.; Holguin, F.; Hong, Y.; Luepker, R. V.; Mittleman, M. A.; et al. Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American heart association. *Circulation* **2010**, *121* (21), 2331−2378.

(2) Heal, M. R.; Kumar, P.; Harrison, R. M. Particles, air quality, policy and health. *Chem. Soc. Rev.* **2012**, *41* (19), 6606−6630.

(3) Oberdörster, G.; Oberdörster, E.; Oberdörster, J. Nanotoxicology: An emerging discipline evolving from studies of ultrafine particles. *Environ. Health Perspect.* **2005**, *113* (7), 823−839.

(4) Kumar, S.; Verma, M. K.; Srivastava, A. K. Ultrafine particles in urban ambient air and their health perspectives. *Rev. Environ. Health* **2013**, *28* (2−3), 117−128.

(5) HEI Review Panel. Understanding the Health Effects of Ambient Ultrafine Particles http://pubs.healtheffects.org/view.php?id=394 (accessed August 20, 2015).

(6) Hoek, G.; Beelen, R.; de Hoogh, K.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ.* **2008**, *42* (33), 7561−7578.

(7) Eeftens, M.; Beelen, R.; De Hoogh, K.; Bellander, T.; Cesaroni, G.; Cirach, M.; Declercq, C.; Dedele, A.; Dons, E.; De Nazelle, A.; et al. Development of land use regression models for PM2.5, PM 2.5 absorbance, PM10 and PMcoarse in 20 European study areas; Results of the ESCAPE project. *Environ. Sci. Technol.* **2012**, *46* (20), 11195−11205.

(8) Hoek, G.; Beelen, R.; Kos, G.; Dijkema, M.; Zee, S. C.; Van Der; Fischer, P. H.; Brunekreef, B. Land use regression model for ultrafine particles in Amsterdam. *Environ. Sci. Technol.* **2011**, *45* (2), 622−628.

(9) Eeftens, M.; Meier, R.; Schindler, C.; Aguilera, I.; Phuleria, H.; Ineichen, A.; Davey, M.; Ducret-Stich, R.; Keidel, D.; Probst-Hensch, N.; et al. Development of land use regression models for nitrogen dioxide, ultrafine particles, lung deposited surface area, and four other markers of particulate matter pollution in the Swiss SAPALDIA regions. *Environ. Health* **2016**, *15* (1), 53.

(10) Rivera, M.; Basagaña, X.; Aguilera, I.; Agis, D.; Bouso, L.; Foraster, M.; Medina-Ramón, M.; Pey, J.; Künzli, N.; Hoek, G. Spatial distribution of ultrafine particles in urban settings: A land use regression model. *Atmos. Environ.* **2012**, *54*, 657−666.

(11) Abernethy, R. C.; Allen, R. W.; McKendry, I. G.; Brauer, M. A land use regression model for ultrafine particles in Vancouver, Canada. *Environ. Sci. Technol.* **2013**, *47* (10), 5217−5225.

(12) Montagne, D. R.; Hoek, G.; Klompmaker, J. O.; Wang, M.; Meliefste, K.; Brunekreef, B. Land Use Regression Models for Ultrafine Particles and Black Carbon Based on Short-Term Monitoring Predict Past Spatial Variation. *Environ. Sci. Technol.* **2015**, *49* (14), 8712−8720.

(13) Saraswat, A.; Apte, J. S.; Kandlikar, M.; Brauer, M.; Henderson, S. B.; Marshall, J. D. Spatiotemporal land use regression models of fine, ultrafine, and black carbon particulate matter in New Delhi, India. *Environ. Sci. Technol.* **2013**, *47* (22), 12903−12911.

(14) Ragettli, M. S.; Ducret-Stich, R. E.; Foraster, M.; Morelli, X.; Aguilera, I.; Basagaña, X.; Corradi, E.; Ineichen, A.; Tsai, M. Y.; Probst-Hensch, N.; et al. Spatio-temporal variation of urban ultrafine particle number concentrations. *Atmos. Environ.* **2014**, *96*, 275−283.

(15) Kerckhoffs, J.; Hoek, G.; Messier, K. P.; Brunekreef, B.; Meliefste, K.; Klompmaker, J. O.; Vermeulen, R. Comparison of Ultrafine Particle and Black Carbon Concentration Predictions from a Mobile and Short-Term Stationary Land-Use Regression Model. *Environ. Sci. Technol.* **2016**, *50* (23), 12894−12902.

(16) Sabaliauskas, K.; Jeong, C. H.; Yao, X.; Reali, C.; Sun, T.; Evans, G. J. Development of a land-use regression model for ultrafine particles in Toronto, Canada. *Atmos. Environ.* **2015**, *110*, 84−92.

(17) Weichenthal, S.; Ryswyk, K.; Van Goldstein, A.; Bagg, S.; Shekkarizfard, M.; Hatzopoulou, M. A land use regression model for ambient ultrafine particles in Montreal, Canada: A comparison of linear regression and a machine learning approach. *Environ. Res.* **2016**, *146*, 65−72.

(18) Hankey, S.; Marshall, J. D. Land Use Regression Models of On-Road Particulate Air Pollution (Particle Number, Black Carbon, PM2.5, Particle Size) Using Mobile Monitoring. *Environ. Sci. Technol.* **2015**, *49* (15), 9194−9202.

(19) Gulliver, J.; De Hoogh, K.; Hansell, A.; Vienneau, D. Development and back-extrapolation of NO2 land use regression models for historic exposure assessment in Great Britain. *Environ. Sci. Technol.* **2013**, *47* (14), 7804−7811.

(20) Wang, M.; Brunekreef, B.; Gehring, U.; Szpiro, A.; Hoek, G.; Beelen, R. A New Technique for Evaluating Land-use Regression Models and Their Impact on Health Effect Estimates. *Epidemiology* **2016**, *27* (1), 51−56.

(21) Klompmaker, J. O.; Montagne, D. R.; Meliefste, K.; Hoek, G.; Brunekreef, B. Spatial variation of ultrafine particles and black carbon in two cities: Results from a short-term measurement campaign. *Sci. Total Environ.* **2015**, *508*, 266−275.

(22) Asbach, C.; Kaminski, H.; Von Barany, D.; Kuhlbusch, T. A. J.; Monz, C.; Dziurowitz, N.; Pelzer, J.; Vossen, K.; Berlin, K.; Dietrich, S.; et al. Comparability of portable nanoparticle exposure monitors. *Ann. Occup. Hyg.* **2012**, *56* (5), 606−621.

(23) Meier, R.; Clark, K.; Riediker, M. Comparative Testing of a Miniature Diffusion Size Classifier to Assess Airborne Ultrafine Particles Under Field Conditions. *Aerosol Sci. Technol.* **2013**, *47* (1), 22−28.

(24) Eeftens, M.; Tsai, M. Y.; Ampe, C.; Anwander, B.; Beelen, R.; Bellander, T.; Cesaroni, G.; Cirach, M.; Cyrys, J.; de Hoogh, K.; et al. Spatial variation of PM2.5, PM10, PM2.5 absorbance and PMcoarse concentrations between and within 20 European study areas and the relationship with NO2 - Results of the ESCAPE project. *Atmos. Environ.* **2012**, *62*, 303−317.

(25) Hsu, H. H.; Adamkiewicz, G.; Houseman, E. A.; Zarubiak, D.; Spengler, J. D.; Levy, J. I. Contributions of aircraft arrivals and departures to ultrafine particle counts near Los Angeles International Airport. *Sci. Total Environ.* **2013**, *444*, 347−355.

(26) Hudda, N.; Gould, T.; Hartin, K.; Larson, T. V.; Fruin, S. A. Emissions from an international airport increase particle number concentrations 4-fold at 10 km downwind. *Environ. Sci. Technol.* **2014**, *48* (12), 6628−6635.

(27) Vert, C.; Meliefste, K.; Hoek, G. Outdoor ultrafine particle concentrations in front of fast food restaurants. *J. Exposure Sci. Environ. Epidemiol.* **2016**, *26* (April), 1−7.

(28) Vineis, P.; Chadeau-Hyam, M.; Gmuender, H.; Gulliver, J.; Herceg, Z.; Kleinjans, J.; Kogevinas, M.; Kyrtopoulos, S.; Nieuwenhuijsen, M.; Phillips, D.; et al. The exposome in practice: Design of the EXPOsOMICS project. *Int. J. Hyg. Environ. Health* **2016**, DOI: 10.1016/j.ijheh.2016.08.001.

(29) Overpass. Overpass Turbo https://overpass-turbo.eu/ (accessed January 18, 2017).

(30) QGIS Development Team. QGIS Geographic Information System. Open Source Geospatial Foundation Project 2016.

(31) R Core Team. Computational Many-Particle Physics. *R Foundation for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, 2008; p 2673.

DOI: 10.1021/acs.est.6b05920
Environ. Sci. Technol. 2017, 51, 3336−3345

(32) Cyrys, J.; Eeftens, M.; Heinrich, J.; Ampe, C.; Armengaud, A.; Beelen, R.; Bellander, T.; Beregszaszi, T.; Birk, M.; Cesaroni, G.; et al. Variation of NO2 and NOx concentrations between and within 36 European study areas: Results from the ESCAPE study. *Atmos. Environ.* **2012**, *62*, 374−390.

(33) Bechle, M. J.; Millet, D. B.; Marshall, J. D. National Spatiotemporal Exposure Surface for NO_2: Monthly Scaling of a Satellite-Derived Land-Use Regression, 2000−2010. *Environ. Sci. Technol.* **2015**, *49* (20), 12297−12305.

(34) Wang, M.; Beelen, R.; Bellander, T.; Birk, M.; Cesaroni, G.; Cirach, M.; Cyrys, J.; de Hoogh, K.; Declercq, C.; Dimakopoulou, K.; et al. Performance of multi-city land use regression models for nitrogen dioxide and fine particles. *Environ. Health Perspect.* **2014**, *122* (8), 843−849.

(35) Basagaña, X.; Rivera, M.; Aguilera, I.; Agis, D.; Bouso, L.; Elosua, R.; Foraster, M.; de Nazelle, A.; Nieuwenhuijsen, M.; Vila, J.; et al. Effect of the number of measurement sites on land use regression models in estimating local air pollution. *Atmos. Environ.* **2012**, *54*, 634−642.

DOI: 10.1021/acs.est.6b05920
Environ. Sci. Technol. 2017, 51, 3336−3345